nature medicine

HIV evolution: CTL escape mutation and reversion after transmission

A J Leslie[1,12], K J Pfafferott[1,12], P Chetty[2], R Draenert[3], M M Addo[3], M Feeney[3], Y Tang[3], E C Holmes[4], T Allen[3], J G Prado[5], M Altfeld[3], C Brander[3], C Dixon[1], D Ramduth[2], P Jeena[2], S A Thomas[2], A St John[6], T A Roach[6], B Kupfer[7], G Luzzi[8], A Edwards[9], G Taylor[10], H Lyall[10], G Tudor-Williams[10], V Novelli[11], J Martinez-Picado[5], P Kiepiela[2], B D Walker[3] & P J R Goulder[1,3]

Within-patient HIV evolution reflects the strong selection pressure driving viral escape from cytotoxic T-lymphocyte (CTL) recognition. Whether this intrapatient accumulation of escape mutations translates into HIV evolution at the population level has not been evaluated. We studied over 300 patients drawn from the B- and C-clade epidemics, focusing on human leukocyte antigen (*HLA*) alleles *HLA-B57* and *HLA-B5801*, which are associated with long-term HIV control and are therefore likely to exert strong selection pressure on the virus. The CTL response dominating acute infection in *HLA-B57/5801*-positive subjects drove positive selection of an escape mutation that reverted to wild-type after transmission to *HLA-B57/5801*-negative individuals. A second escape mutation within the epitope, by contrast, was maintained after transmission. These data show that the process of accumulation of escape mutations within HIV is not inevitable. Complex epitope- and residue-specific selection forces, including CTL-mediated positive selection pressure and virus-mediated purifying selection, operate in tandem to shape HIV evolution at the population level.

Evasion of the host CTL response through mutation of key epitopes is a major challenge to natural or vaccine-induced immune control of HIV[1–4]. This phenomenon of CTL escape mutation in immunodeficiency virus infection has been well documented[1–10]. Recent studies suggest that, through escape mutation, CTL has a major role in driving evolution of HIV[5,9,10]. However, the extent to which this actually occurs remains to be determined. It is clear that CTL escape mutations are common, and are selected by CTL responses restricted by a wide array of different HLA molecules[9]. However, the question of whether these mutations are transmissible and then stable in the absence of the *HLA* allele that drove their selection has yet to be addressed. These two parameters need to be established to quantify the role of the CTL response in driving HIV evolution at a population level. It is also of vital importance in terms of vaccine design to determine which immune responses will remain relevant and which will be irrevocably lost through escape mutations.

To address this question, we sought examples of *HLA-B57*- and *HLA-5801*-restricted CTL escape, and observed the outcome of transmission of these mutations to *HLA-B57/5801*-negative individuals. These *HLA* alleles were chosen because *HLA-B57* and the closely related *HLA-B5801* are associated with effective control of HIV[11–15] (P.J.R.G. *et al.*, unpublished data). We therefore hypothesized that the *HLA-B57/5801*-restricted CTL response represents a strong selection force that would be associated with escape mutation. We focused here on the Gag polyprotein, as this is a highly immunogenic region of HIV-1 (refs. 16–18).

TSTLQEQIAW variation is associated with *HLA-B57/5801*

Gag sequences from 311 subjects with chronic HIV infection were analyzed. Variation in the epitope TSTLQEQIAW (designated TW10; Gag HXB2 residues 240–249) was observed in association with expression of *HLA-B57* or *HLA-B5801* (**Table 1a**). TW10 dominates the CTL response in acute HIV infection in *HLA-B57* individuals[19], and is presented by the closely related *HLA-B5801* allele[14]. The dominant change in TW10 is a substitution of Thr by Asn at residue 242 (T242N). This mutation occurs in C-clade infections, arising in 84% and 63% of *HLA-B57*-positive and *HLA-B5801*-positive subjects, respectively, and in 0% of *B57/5801*-negative subjects ($P < 0.0001$ in each case). Data from B-clade-infected subjects reveal a similar picture (**Table 1a**). Again, the T242N mutation is highly prevalent in *HLA-B57*-posi-

[1]Department of Pediatrics, Fuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 3SY, UK. [2]HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of Natal, Congella 4013, Durban, South Africa. [3]Partners AIDS Research Center and Harvard Medical School, Boston, Massachusetts 02129, USA. [4]Department of Zoology, University of Oxford, Oxford OX1 3SY, UK. [5]IrsiCaixa Foundation, University Hospital 'Germans Trias i Pujol', 08916 Barcelona, Spain. [6]Queen Elizabeth II Hospital, Bridgetown, Barbados. [7]University of Bonn, 53105 Bonn, Germany. [8]Department of Genitourinary Medicine, High Wycombe General Hospital, HP11 2TT, UK. [9]The Harrison Clinic, Radcliffe Infirmary Hospital, Oxford, OX2 6HE, UK. [10]Imperial College School of Medicine, St Mary's Hospital, London W2 1NY, UK. [11]Great Ormond Street Hospital for Sick Children, London WC1N 3JH, UK. [12]These authors contributed equally to this work. Correspondence should be addressed to P.J.R.G. (philip.goulder@ndm.ox.ac.uk).

Published online 8 February 2004; doi:10.1038/nm992

tive and *HLA-B5801*-positive subjects (79% and 93%, respectively), and is not found in any *B57/5801*-negative individuals ($P < 0.0001$ in each case).

An interclade difference exists within the TW10 epitope at residue 248, where the consensus is Gly (G) in B-clade strains and Ala (A) in C-clade strains. In C-clade-infected subjects, variation at this position is weakly associated with *HLA-B57* only ($P = 0.03$). By contrast, there is a high degree of variation at this position in B-clade-infected subjects, predominantly involving the G248A change, again strongly associated with *HLA-B57* ($P < 0.0001$) but not *HLA-B5801*. Unlike T242N, G248X is found in substantial numbers in *HLA-B57/5801*-negative individuals (**Table 1a**). Together, these results indicate that TW10 may be under immune selection pressure that results predominantly in a T242N substitution in both B-clade and C-clade infection, as well as an additional G248X substitution, particularly in B-clade infections.

TW10 variation arises after transmission

To show that the TW10 variants arise after transmission to *HLA-B57/5801*-positive subjects, we sought examples of vertical and horizontal transmission in which *HLA-B57/5801*-negative donors, infected with wild-type virus, transmit to subjects expressing *HLA-B57* or *HLA-B5801*. Six such mother-child transmission pairs were identified, with viral sequences obtained from all six mothers encoding wild-type TW10 and from five of the six *HLA-B57/5801*-positive children carrying the T242N and/or G248A mutations (**Table 1b**). In the two adult pairs in which the direction of horizontal transmission was unambiguous, only the *HLA-B57/5801*-positive subjects possessed the T242N mutation. Likewise, in three additional adult pairs in which the direction of horizontal transmission was ambiguous, only virus isolated from the *HLA-B57/5801*-positive subjects possessed the T242N mutation (**Table 1c**). In this last group, the data could indicate either transmission of wild-type virus to the *B57/5801*-positive individual followed by escape, or transmission of the T242N mutant virus followed by reversion to wild type in the *HLA-B57/5801*-negative recipient. The authenticity of all of these transmission pairs was confirmed by phylogenetic analysis (**Supplementary Fig. 1** online).

Table 1 Variation in TW10

a

	HLA-B57			HLA-B5801			Other		
C-clade ($n = 183$)	TSTLQEQIAW	n	%	TSTLQEQIAW	n	%	TSTLQEQIAW	n	%
	--N-------	17	53	--N-------	24	56	----------	90	83
	--N-----T-	5	16	----------	8	19	--------T-	7	6
	--N-A-----	3	9	----P-----	3	7	--S-------	3	3
	--N-----V-	2	6	--S-------	2	5	--------V-	3	3
	-------M--	1	3	----A-----	2	5	--------Q-	1	1
	-------MT-	1	3	--N-----N-	1	2	--------VD-	1	1
	--------I-	1	3	--N-----T-	1	2	-------VT-	1	1
	--------Q-	1	3	--N-----G-	1	2	-----D----	1	1
	-------V--	1	3	---------Q-	1	2	--------G-	1	1
	Total variation	100% ($P < 0.0001$)		Total variation	81% ($P < 0.0001$)		Total variation	17%	
	T242N	84% ($P < 0.0001$)		T242N	63% ($P < 0.0001$)		T242N	0%	
	A248X	25% ($P = 0.03$)		A248X	9% (NS)		A248X	10%	
B-clade ($n = 128$)	TSTLQEQIGW	n	%	TSTLQEQIGW	n	%	TSTLQEQIGW	n	%
	--N-----A-	14	41	--N-------	8	53	----------	55	70
	--N-------	9	26	--N-----A-	5	3	--------A-	14	18
	--------A-	4	12	--N-----T-	1	7	--S-------	2	3
	--N-----Q-	2	6	--------A-	1	7	--------T-	2	3
	--N-A---A-	1	3				-------VT-	1	1
	--------D-	1	3				----------E	1	1
	---B-----A-	1	3				--S-----A-	1	1
	--N-----T-	1	3				---P------	1	1
	----------	1	3				-------VA-	1	1
							----A---A-	1	1
	Total variation	97% ($P < 0.0001$)		Total variation	100% ($P < 0.0001$)		Total variation	30%	
	T242N	79% ($P < 0.0001$)		T242N	93% ($P < 0.0001$)		T242N	0%	
	G248X	71% ($P < 0.0001$)		G248X	47% (NS)		G248X	27%	

b

Transmission pair	HLA type	Route	Time point	TSTLQEQIGW
PF750214-Mother	B42/44	Vertical	2–3 years pp[a]	----------
990918KF-Child	B57/42		2 years pp	--------A-
			3 years pp	--N-----A-
BS660508-Mother	B42/71	Vertical	2 years pp	----------
000401CM-Child	B57/71			---N-----A-
DF680505-Mother	B7/53	Vertical	6 years pp	----------
951111MF-Child	B5801/53			---------A-
Don-17	B44/49	Horizontal	NK[b]	----------
Don-23	B57/08			--N-------
Bonn-F	B7/–	Blood-borne	11 years[c]	----------
Bonn-G	B57/27			--N-------
				TSTLQEQIAW
J120-Mother	B42/81	Vertical	6 years pp	----------
J120-Child	B57/42			--N------T-
J143-Mother	B44/–	Vertical	4 years pp	----------
J143-Child	B5801/44			--N-------
J3-Mother	B53/71	Vertical	NK	----------
J3-Child	B5801/71			----------

c

Transmission pair	HLA type	Route	Genetic material	TSTLQEQIGW	Number of clones
043M-Wife	B35/49	Horizontal	DNA	----------	Population sequencing
043D-Husband	B5801/27			--N-------	
068M-Wife	B27/53	Horizontal	DNA	----------	7/7
			RNA	----------	5/5
068D-Husband	B5801/72		DNA	--N-------	2/2
			RNA	--N-------	3/4
				--N-----E-	1/4
6008-Wife	B27/44	Horizontal	DNA	--S-------	Population sequencing
6007-Husband	B5801/53			--N-------	

(a) All sequences are from HIV antibody–positive individuals, and represent population sequences from proviral DNA. NS, not significant. (b) Transmission of wild-type TW10 to *B57/5801*-positive subjects in unambiguous direction. J120, J143 and J3 mother-and-child pairs were infected with C-clade HIV; remaining transmission pairs were infected with B-clade HIV. Don-17 and Don-23 are a previously reported horizontal transmission pair[44], in which patient history indicates that Don-17 infected Don-23. Bonn-F and Bonn-G, also previously reported[45], are hemophiliacs infected with the same batch of contaminated factor VIII. All sequence data shown represent population sequences from proviral DNA. [a]Postpartum. [b]Time after transmission not known. [c]Time after transmission. B57 and B5801 are shown in bold. (c) Transmission of HIV within husband-wife pairs in which the direction of transmission was unknown. All subjects were infected with B-clade virus; time of transmission is unknown.

T242N arises as a result of positive selection

To determine whether TW10 is under positive selection from *HLA-B57/5801*-restricted CTLs, we examined selection pressures using a maximum-likelihood method[20]. This method revealed nine sites under positive selection when *B57/5801*-positive and non-*B57/5801* subjects were grouped together (**Supplementary Table 1** online), including residues 242 (mean ratio of nonsynonymous to synonymous nucleotide substitutions (d_N/d_S) = 3.216)

and 146 (mean d_N/d_S = 3.215). Both results were highly significant, with Bayesian posterior probabilities of 1.000 and 0.999, respectively. When *HLA-B57/5801*-positive subjects were excluded from the analysis, only residues 242 and 146 were no longer identified as being under positive selection, suggesting that selection at these sites is operating through the *HLA-B57/5801* alleles. The same analysis conducted using the B-clade sequence data similarly identified residues 242 and 146 as being under positive selection (mean d_N/d_S = 2.34 and 2.23, respectively).

TW10 variants are *B57/5801* CTL escape mutants

The optimal epitope TW10 (TSTLQEQIGW) was previously defined in B-clade-infected subjects expressing *HLA-B57* or *B5801* (ref. 14). TW10 (TSTLQEQIAW) was confirmed as the optimal epitope in C-clade infection (**Fig. 1a**). Recognition of the commonly occurring variants at residues 242 and 248 was evaluated in *B57*- and *B5801*-positive subjects with B- or C-clade infection, respectively (**Fig. 1b–e**). The pattern of recognition was similar in each case, with reduced recognition equivalent to 1–2 logs peptide concentration of variants with single amino-acid substitutions, and complete abrogation of a response to peptides bearing double mutations. Thus, these *HLA-B57*- and *B5801*-associated mutations within TW10, which dominate the viral population in chronically infected individuals expressing these alleles, represent escape mutations.

T242N 'reverts' in the absence of *B57/5801*

We noted above that T242N is seen in 0 of 187 *HLA-B57/5801*-negative subjects with chronic HIV infection. The most likely explanation for this is that T242N mutant viruses are not transmitted, or that the T242N mutant is outcompeted by wild-type virus ('reverts') after transmission to *B57/5801*-negative individuals. To determine whether T242N is transmitted, instances of transmission from *HLA-B57/5801*-positive subjects to individuals lacking these alleles were sought. In one such case of intrapartum mother-to-child transmission (**Fig. 2a**), the B-clade-infected mother (SMH-05M; *HLA-B57*-positive) carried both the T242N and G248A mutations (46 of 46 clones) at 5 months, 7 months and 8 years postpartum, indicating that both mutations are stable in a *B57/5801*-positive individual for this duration. In the child (SMH-05C; *HLA-B7* homozygous) at 2 months postpartum, all clones possessed the T242N and G248A mutations, confirming that both mutations were transmitted. By 5–7 months of age, the frequency of the T242N mutation declined to ~50%, and by 5 years none of the 16 clones isolated from plasma RNA possessed the T242N mutation. (**Table 2** and **Fig. 2a**). Although 3 of 17 clones isolated from proviral DNA at 8 years of age encoded T242N, phylogenetic analysis confirmed that these are closely related to the transmitted virus, and thus represent archival sequences (**Fig. 2b**). In contrast to T242N, the G248A mutation was present in all SMH-05C clones from all time points, suggesting that G248A is stable (does not revert) in the absence of the selecting *HLA* allele.

A second mother-to-child transmission pair, 997-Mother (*HLA-B18/57*) and 997-Child (*HLA-B18/42*), was identified (**Table 2** and **Fig. 2c**). This pair was C-clade infected. T242N was present in the mother antenatally and at all time points postpartum, as expected. In the child, who was infected 3–9 months postpartum, 0 of 22 clones sequenced from RNA had the T242N mutation at the 9-month time point. These data again support the hypothesis that the T242N mutation reverts in the absence of *HLA-B57/5801*.

Analysis of 187 chronically infected *HLA-B57/5801*-negative individuals revealed none with virus expressing the T242N mutation. In contrast, virus expressing the T242N mutation was isolated from 4 of 19 *HLA-B57/5801*-negative subjects with acute HIV infection (**Table 2**). This further demonstrates that T242N is transmitted horizontally, and is in itself evidence that T242N reverts at some point between acute and chronic infection in *HLA-57/5801*-negative individuals in whom the transmitted virus carries the T242N mutation.

One of these acutely infected individuals, AC-33, was studied longitudinally, revealing that T242N also reverts over time in this situation (**Table 2** and **Fig. 2d**). This subject was placed on antiretroviral therapy (ART) immediately after diagnosis of HIV infection. During ART, viral replication was suppressed to undetectable levels (<50 copies/ml). One might thus anticipate that T242N reversion would take longer in this setting than in patient SMH-05C, who did not receive ART until 4 years after infection, or in patient 997-C, who never received ART.

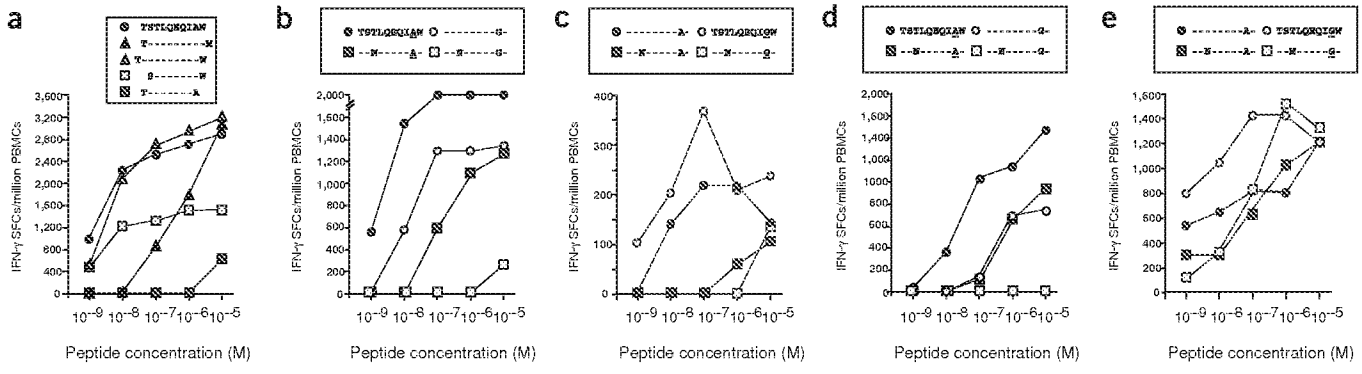

Figure 1 TW10 variants are escape mutations. IFN-γ production was recorded as spot-forming cells (SFCs) per million PMBCs. Results shown are representative of assays using PBMCs from seven subjects expressing either *HLA-B57* or *HLA-B5801*. (**a**) Recognition of TW10 and suboptimal 9-mer and 11-mer overlapping peptides using PBMCs from C-clade-infected donor A-005-M (*HLAB57/7*). (**b–d**) TW10 variant recognition. (**b**) *HLA-B57*-positive, C-clade-infected subject A-005-M (**c**), *HLA-B57*-positive, B-clade-infected subject SWS (*HLA-B57*; **d**), *HLA-B5801*-positive, C-clade-infected subject PS-032-M (*B5801*; **e**) and *HLA-B5801*-positive, B-clade-infected subject, 90-92851-RI (*B5801*). Sequence of wild-type TW10 is shown in full in each key.

Reversion revealed by *B57/5801* 'footprints'

To further substantiate the hypothesis that T242N undergoes reversion in the absence of *HLA-B57/5801*, we next sought T242N-linked mutations that are stable in the absence of *B57/5801*. The term 'footprints'[21] was previously coined to describe sequence polymorphisms associated with particular *HLA* alleles. Thus, the presence of characteristic polymorphisms can indicate the impact of specific *HLA* alleles on HIV. In *HLA-B57/5801*-negative individuals, the association of such footprints of the T242N mutation with variation in TW10, which we have shown is strongly associated with *HLA-B57/5801*, would support the hypothesis that the ancestral virus in these subjects originated in *HLA-B57/5801*-positive individuals (**Fig. 3**). Two sites were identified at which variation from the consensus sequence was associated with T242N in the *HLA-B57/5801*-positive group: H219X (X = Q, P or R; $P < 0.0001$) and A146X (X = P, T, S, V, N or H; $P = 0.0008$). In the case of A146P, this polymorphism represents a processing escape mutation (R.D. *et al.*, unpublished data). In the case of H219X, no *HLA-B57* or *B5801*-restricted epitope has been identified for this region[14,18,22], and there is no potential *B57* or *B5801*-restricted epitope that would satisfy the defined *HLA-B57/5801*-binding motif[23].

Table 2 Fate of TW10 variants after transmission

Subject	HLA type	Time point	Genetic material	TSTLQEQIGW	Number of clones
SMH-05-Mother	**B57/7**	5 months	DNA	--N-----A-	16/16
		7 months	RNA	--N-----A-	17/17
		8 years	DNA	--N-----A-	13/13
			RNA	--N-----A-	Population sequencing
SMH-05-Child	**B7/–**	2 months	DNA	--N-----A-	14/14
		5 months	DNA	--N-----A- --------A-	12/21 9/21
		7 months	RNA	--N-----A- --------A-	6/12 6/12
		5 years	RNA	--------A-	16/16
		8 years	DNA	--N-----A- --------A-	3/17 14/17
				TSTLQEQIAW	
997-Mother	**B57/18**	–8 weeks	DNA	--N-------	Population sequencing
		1 week	DNA	--N-------	Population sequencing
		6 weeks	DNA	--N-------	Population sequencing
		3 months	DNA	--N-------	Population sequencing
			RNA	--N-------	15/15
		9 months	DNA	--N-------	Population sequencing
997-Child	**B42/18**	9 months	RNA	-------M--	22/22
				TSTLQEQIGW	
AC-04 AC-27 AC-14	**B18/44** **B8/35** **B8/62**	Pre-sero	DNA	--N-----A- --N------- --N-----A-	Population sequencing
AC-33	**B7/44**	Day 0 Pre-sero	RNA	--N------- --N----M--	6/7 1/7
		Day 43	DNA	--N-------	8/8
		Day 818	RNA	--N------- ----------	2/7 5/7
		Day 983	RNA	--N------- ----------	3/14 11/14
		Day 1541	RNA	----------	4/4

SMH-05-Mother/SMH-05-Child and 997-Mother/997-Child are vertical transmission pairs in which *HLA-B57*-positive mothers (infected with B-clade and C-clade virus, respectively) transmitted virus to their *HLA-B57/5801*-negative children. DNA, sequence obtained from proviral DNA; RNA, sequence obtained from viral RNA. All time points are postpartum, apart from one time point (–8 weeks) at which the 997-Mother sample was antenatal. AC-04, AC-27, AC-14 and AC-33 are *HLA-B57/5801*-negative adults acutely infected with B-clade virus possessing T242N. All were treated with ART in acute infection. Subject AC-33 was followed longitudinally, with time points given as days since seroconversion. Pre-sero, before seroconversion. B57 and B5801 are shown in bold.

The H219X variant appears subsequent to the development of the T242N mutation. In *B57/5801*-positive subjects, H219X is seen only in the presence of T242N (32 of 32 instances; **Table 3a**), compared with 64 cases where T242N is found in the absence of H219X and 28 cases where neither variant is found ($P > 0.0001$). In addition, longitudinal sequencing of virus in SMH-05M demonstrated the acquisition of this H219X variant after fixation of the T242N mutation (**Table 3b**). The hypothesis that the H219 variant may be maintained after transmission, whereas the T242N variant reverts, is also supported by analysis of two husband-wife pairs (6007/6008 and 068D/068M; **Table 3b**). In both cases, the T242N mutation present in the putative donors (6007 and 068D; both *HLA-B5801*-positive) was linked with the H219Q variant, whereas in the spouses (6008 and 068M; both *HLA-B5801*-negative), only the footprints of these variants remain.

We next examined the *B57/5801*-negative subjects to test the hypothesis that the T242N footprints H219X and A146X are associated with TW10 mutations other than T242N, as T242N would be expected to revert in the absence of *HLA-B57/5801*. We found that both H219X and A146X were associated with variation in TW10 ($P = 0.0009$ and $P = 0.0002$, respectively), in spite of the absence of T242N from these subjects. These data, which show an association between the two identified T242N footprints, H219X and A146X, and TW10 variation within *HLA-B57/5801*-negative subjects, support the hypothesis that the ancestral virus in these individuals originated from *HLA-B57/5801*-positive subjects in which the T242N variant was present. Taken together, these data provide compelling evidence that the T242N mutation is transmitted to, and subsequently reverts in, *B57/5801*-negative subjects.

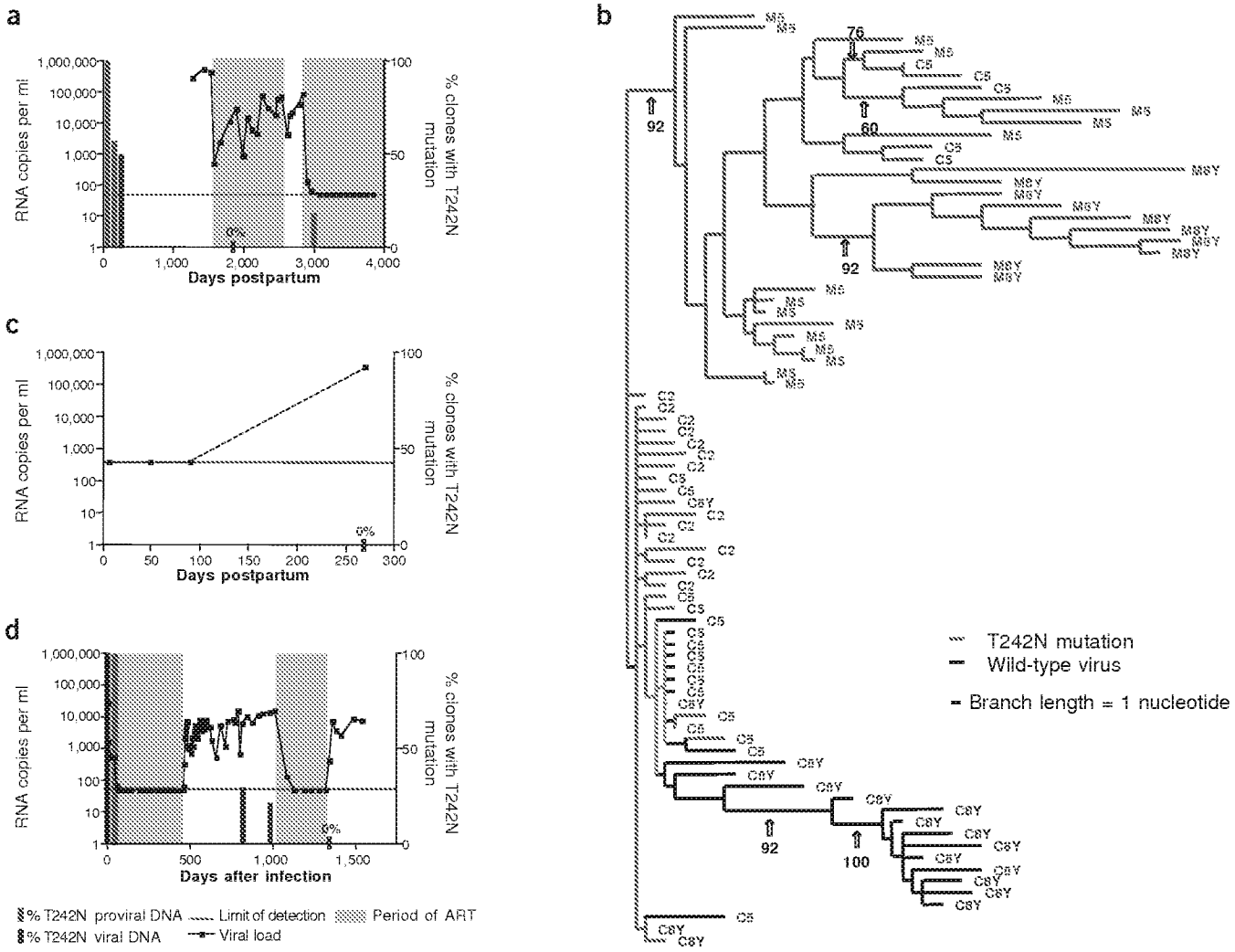

Figure 2 Reversion of TW10 variants after transmission to *HLA-B57/5801*-negative subjects. (**a**,**c**,**d**) Viral loads of three subjects, SMH-05C (**a**), 997C (**c**) and AC-33 (**d**), showing intrapartum mother-to-child, postpartum mother-to-child and horizontal transmission, respectively. (**b**) Maximum-likelihood phylogenetic tree of SMH-05-Mother and SMH-05-Child proviral DNA clones. Blue, SMH-05-Child clones from 2 months (C2), 5 months (C5) and 8 years (C8Y); pink, SMH-05-Mother clones from 5 months (M5) and 8 years (M8Y); T242N mutation is mapped onto the tree. Lineages in which ancestral state is equivocal are highlighted in gray. Yellow arrows indicate major sublineages supported by high bootstrap values.

DISCUSSION

These data provide the first clear evidence of reversion after transmission of an HIV-1 CTL epitope in which escape has occurred in the donor. We show that T242N is an escape mutation in the TW10 epitope that is selected in the majority of individuals possessing either *HLA-B57* or *HLA-5801*. *B57/5801*-positive individuals can transmit the T242N mutant, and such transmission has occurred in numerous instances. The T242N mutation was found in 0 of 187 of *HLA-B57/5801*-negative individuals with chronic HIV infection, compared with 4 of 19 individuals in the acute phase of infection ($P = 0.00005$), providing strong evidence that T242N is transmitted and, in the absence of *HLA-B57/5801*, reverts to wild type after acute infection. We present longitudinal data from three *HLA-B57/5801*-negative subjects to demonstrate that the transmitted T242N variant is replaced by wild-type virus over time. Further evidence of T242N transmission, followed by reversion, derives from the analysis of genetic footprints that are linked with T242N in *HLA-B57/5801*-positive subjects, and that can be identified in *B57/5801*-negative subjects in whom T242N is no longer present.

A previous study examining escape mutations in the HLA-B27-restricted epitope KK10, a second epitope associated with long-term control of HIV infection, indicated that the occurrence of reversion after transmission in this instance is rare[10]. Intrapatient reversion after progression to AIDS is also rare, having been suggested by only a single clone out of >150 clones analyzed from that study and others[10,24]. Thus, the B27-KK10 escape variants generally remain stable in the absence of the evolutionary pressure that originally selected them[10,24]. This contrasts with T242N, which consistently reverts after transmission to a *B57/5801*-negative individual. This distinction highlights the importance of determining the fate of CTL escape mutations. If they are transmissible and stable after transmission, they will spread through the population and the epitope will eventually become extinct. Conversely, if the escape mutant reverts, the epitope will remain useful for CTL-based vaccine design.

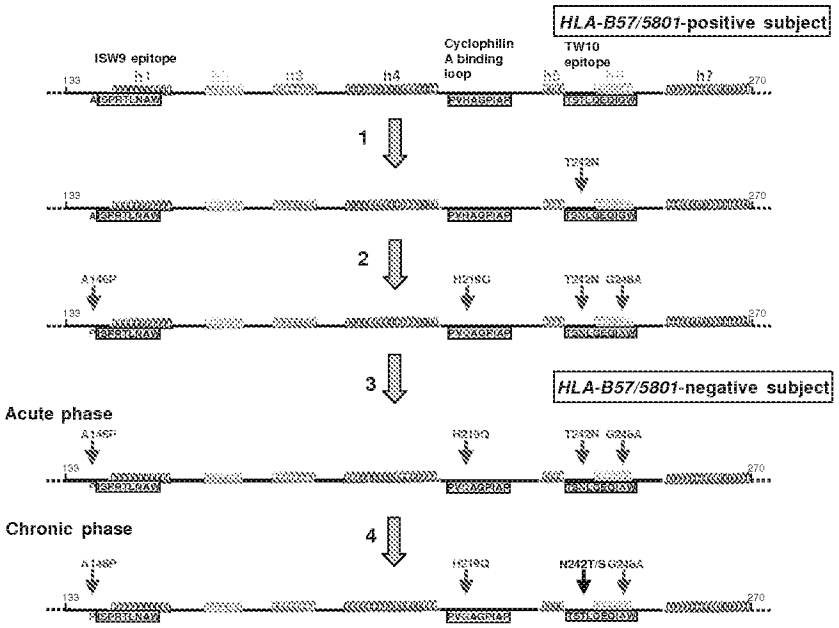

Figure 3 Patterns of *HLA-B57/5801*-associated mutation and reversion before and after transmission. (1) Transmission of wild-type virus to *HLA-B57*- or *HLA-B5801*-positive subject, in whom the T242N mutation arises early. (2) Accumulation of *HLA-B57/5801*- and T242- linked mutations. (3) Transmission of *HLA-B57/5801*-adapted virus to *B57/5801*-negative individual. (4) T242N mutation reverts, whereas stable *B57/5801*- and T242N-linked mutations remain as footprints of the *HLA-B57/5801* allele. Shown are structural features of N-terminal portion (residues 131–270) of p24 Gag: α-helices 1–7 (h1–h7), cyclophilin A binding loop, and HLA-B57/5801-restricted epitopes ISPRTLNAW (ISW9; Gag residues 147–155) and TSTLQEQIGW (TW10; residues 240–249).

The likely explanation for T242N reversion is that there is an associated fitness cost. Drug-resistance mutations present an obvious analogy, as they typically do not persist in untreated patients, despite conferring greater fitness in the presence of ART, because they impair the replication of wild-type virus[25]. Thus, ART-resistant mutants may pay an 'evolutionary penalty' for their survival[26]. Similarly, the T242N mutation may represent the evolutionary penalty paid by HIV to escape TW10-specific CTLs. TW10 is the immunodominant CTL response in acute infection[19], so even an attenuated virus would have a selection advantage over one that is recognized and eliminated by TW10-specific CTLs. Recent data indicate that in the rare *HLA-B57-*

Table 3 Footprints of HLA-B57 and HLA-B5801

a

HLA-B57/5801-positive

B-clade

A146X	H219X	TSTLQEQIGW	*n* = 49
P	-	--N--------	6
P	-	--N------A-	5
P	Q	--N------A-	4
-	Q	--N------A-	4
-	Q	--N--------	3
-	Q	--N------T-	2
P	Q	--N--------	2
-	Q	--N-----Q-	2
P	-	---------A-	2
P	Q	--N-----A-	1
P	Q	--N-------	1
-	Q	--N-A---A-	1
-	-	--N------A-	5
-	-	--N--------	5
-	-	---------	3
-	-	---------D-	1
-	-	---H-----A-	1
-	-	----------	1

C-clade

A146X	H219X	TSTLQEQIAW	*n* = 75
P	-	--N--------	18
P	Q	--N------T-	4
P	-	--N-A------	3
P	Q	--N--------	2
-	Q	--N--------	2
T	-	--N--------	2
-	Q	--N-----T-	1
H	Q	--N----V--	1
P	P	--N----V--	1
P	R	--N--------	1
P	-	--N------N-	1
P	-	--N------T-	1
P	-	--N------N-	1
P	-	--------MT-	1
P	-	--------M--	1
S	-	----------	1
-	-	--N--------	16
-	-	---P-------	2
-	-	---------Q-	2
-	-	----A------	2
-	-	--S--------	2
-	-	--N------G-	1
-	-	--------V--	1
P	-	----------	1
-	-	----------	7

HLA-B57/5801-negative

B-clade

A146X	H219X	TSTLQEQIGW	*n* = 79
P	-	---------A-	5
P	Q	---------T-	2
-	Q	--------VT-	1
-	Q	--S------A-	1
T	Q	---------A-	1
-	Q	--S--------	1
P	-	--S--------	1
P	-	---A-----A-	1
P	-	----------	4
-	Q	----------	4
T	-	----------	1
V	-	----------	1
-	-	----------	45
-	-	---------A-	8
-	-	---------E-	1
-	-	---P-------	1
-	-	--------VA-	1

C-clade

A146X	H219X	TSTLQEQIAW	*n* = 108
-	Q	----------	6
S	-	--------T-	1
S	-	----------	1
P	Q	--------V--	1
-	Q	--S--------	1
-	Q	---------Q-	1
-	Q	---------T-	1
P	Q	----------	1
-	Q	--------VT-	1
N	-	---------G-	1
-	-	----------	77
-	-	---------T-	5
T	-	----------	3
P	-	----------	2
-	-	--S--------	2
-	-	--------V--	2
-	-	--------VD-	1
-	-	-----D-----	1

b

Subject	Time point	A146X	H219X	TSTLQEQIGW	Genetic material	Number of clones
SMH-05M (B57/07)	5 months pp[a]	P	-	--N------A-	DNA	16/16
	8 years pp	P	Q	--N------A-	DNA	9/13
		P	-	--N------A-		4/13
026-BMC (B57/42)	10 years pp	-	-	----------		2/5
		-	Q	--N------A-	DNA	1/5
		P	Q	--N------A-		2/5
6007 (B5801/53)	NK[b]	-	Q	--N-------	DNA	Population Sequencing
6008 (B27/44)	NK	-	Q	--S-------	DNA	Population Sequencing
068-D (B5801/72)	NK	-	Q	--N-------	DNA	2/2
		-	Q	--N-------	RNA	3/4
		-	Q	--N------E-	RNA	1/4
068-M (B27/53)	NK	-	Q	----------	DNA	7/7
		-	Q	----------	RNA	5/5

(a) H219X and A146X are associated with T242N in the *B57/5801*-positive group. In the *B57/5801*-negative group, H219X and A146X are associated with variations in TW10 other than T242N. (b) Patterns of mutation and reversion of TW10-associated variants. Longitudinal analysis of *HLA-B57*-positive subject SMH-05M shows acquisition of H219Q mutation after fixation of T242N mutation. Clones sequenced from 026-BMC show presence of H219Q only in association with T242N. The two chronically infected husband-wife pairs 6007/6008 and 068D/068M further support the hypothesis that H219Q is maintained, whereas T242N is lost after transmission to a *B57/5801*-negative donor. [a]Postpartum. [b]Time after transmission not known. B57 and B5801 are shown in bold.

positive individuals in whom the TW10 epitope has not mutated, viremia is successfully suppressed for decades to <50 copies/ml (ref. 27). Once selection pressure from TW10-specific CTLs is removed, as in transmission to a non-*B57/5801*-expressing individual, one would expect natural selection to return the virus to its fitter, wild-type state. Even minor enhancements of viral fitness can quickly lead to replacement of the less fit variants in a population[28].

One might further speculate that the early dominance of the TW10-specific response in acute infection, followed by the emergence of escape variants conferring an attenuated replicative state upon the virus, could be causally linked to the association of *HLA-B57/B5801* with long-term suppression of HIV infection. In our ART-naive C-clade-infected study group, *HLA-B57/5801*-expressing subjects whose virus encoded T242N had a median viral load of 10,342 HIV RNA copies per ml plasma, still significantly lower than in the *HLA-B57/5801*-negative group (median viral load = 28,125 HIV RNA copies per ml plasma; $P = 0.02$ by Mann-Whitney test). This T242N mutation may therefore contribute to the control of viremia observed in infected persons expressing *HLA-B57/B5801*, despite the loss of CTL recognition through this TW10 specificity.

The precise mechanism by which the T242N mutation might affect viral fitness is unknown. Gag residue 242 falls within the HIV capsid protein, p24, which has crucial roles in viral function[29] and is a highly conserved region of the HIV genome[30]. Single-point mutations in p24 can have a profound effect on viral fitness[31–34], and linker-insertion mutations between residues 241–242 and 242–243 destroy HIV's ability to replicate[32,35,36]. T242N mutant viruses can persist for at least 8 years, however, replicating and evolving (**Fig. 2b**). Under such circumstances, one might expect HIV to accumulate other mutations that restore viral fitness to some degree, as suggested for the late-developing B27-KK10 escape mutation[24]. The H219X mutation, universally observed in association with T242N in *B57/5801*-positive subjects, may represent a similar compensatory mutation. If so, however, the fact that 22 *B57/5801*-negative subjects possessed H219X in the absence of T242N suggests that, unlike the compensatory changes associated with KK10 escape, H219X does not fully restore viral fitness. H219 lies within the cyclophilin A binding loop, which essential to HIV's life cycle[35–37]. H219 is directly involved in cyclophilin A binding, and the H219Q mutation results in a 4.8-fold decrease in binding affinity[37]. It would thus be unlikely for the H219Q mutation to be selected unless it was linked to some separate fitness advantage—in this case, hypothetically, reducing the negative effect of T242N.

In these studies, we have illustrated two contrasting outcomes of CTL escape mutation. In the case of the T242N mutation, which we hypothesize is associated with a fitness cost to the virus, we observe reversion after transmission to *B57/5801*-negative recipients. Conversely, in the case of the G248A mutation, occurring within the same TW10 epitope, we observe maintenance after transmission to recipients lacking *B57/5801*. The inference that this CTL escape mutation arises at an insignificant fitness cost to the virus is supported by a study that examined this particular G248A mutation and found no effect on viral fitness[34]. Between these two extremes, there is likely to be a broad spectrum of CTL escape mutations whose presence may be associated with greater or lesser degrees of constraint on viral replication. Our findings illustrate that intrapatient evolution of HIV driven by CTL escape does not necessarily translate into evolution of HIV at the population level. The degree to which escape mutations revert or are maintained will determine the extent to which they contribute to the evolution of HIV as it spreads through a population.

In conclusion, we have identified two escape mutations in the immunodominant *HLA-B57/5801*-restricted Gag epitope TW10, one of which undergoes reversion and one of which is stable in the absence of *B57/5801*. This highlights the diverse outcomes of CTL escape mutation, and confirms the need to clarify the fate of escape mutations in order to determine their future relevance to vaccine design. These findings also suggest that the role of CTLs in shaping HIV evolution may not be a simple one, and that intrapatient viral evolution does not necessarily translate to evolution of HIV at the population level. Rather, it is the complex interaction between different selection forces—positive selection pressure from CTLs on one hand, and purifying selection in the virus on the other—that determines the role of CTLs in shaping HIV evolution.

METHODS

Subjects. The 128 C-clade-infected subjects analyzed in this study were collected from Durban, South Africa, and consisted predominantly of Zulu and Xhosa women recruited from the Cato Manor antenatal clinic. All subjects were ART-naive. The median viral load was 14,892 HIV RNA copies per ml plasma (range <400–698,000), and the median absolute CD4 count was 449 cells/mm^3 (range 101–1,215 cells/mm^3). The 97 B-clade-infected subjects were collected from diverse sources encompassing Europe, the Caribbean and North America. Of these subjects, 17% were receiving ART at the time of analysis. The median viral load in those not receiving ART was 24,600 RNA copies per ml plasma (range <400 to >750,000), and the median CD4 count for those not receiving ART was 420 cells/mm^3 (range 131–1,280 cells/mm^3). This study was approved by institutional review boards, and all subjects gave written informed consent.

Sequencing of proviral DNA and viral RNA. We extracted genomic DNA from peripheral blood mononuclear cell (PBMC) pellets using the Puregene DNA isolation kit (Gentra). We then amplified HIV *gag* sequences using a nested PCR as described[38], using the following *gag*-specific primers: 5′-CTAGCAGTGGCGCCCGAACA-3′ and 5′-ACAGTCTTTCATTTGGTGTCCTTC-3′ for first-round outer PCR, and 5′-TCTCTCGACGCAGGACTC-3′ and 5′-TTTCCACATTTCCAACAGCC-3′ for second-round inner PCR. The PCR product was purified by PEG precipitation and either directly sequenced (referred to in the text as 'population sequencing'), or cloned as previously described[38] using a TOPO TA cloning kit (Invitrogen). We recovered viral RNA from plasma samples using the Nucleospin RNA extraction kit (Mackery-Nagel). During this step, we added RNA-free DNase (Qiagen) to ensure removal of proviral DNA. We then synthesized a cDNA library from the RNA using the Reverse-iT 1st Strand Synthesis kit (ABgene), using random oligonucleotides to prime the reaction. We then amplified and sequenced *gag* fragments from the cDNA as described above. All sequencing was done with BigDye Terminator v3.0 Ready Reaction mix (Applied Biosystems), using the two inner PCR primers listed above and four additional primers (5′-CTGCACTATAGGATAATTTTGAC-3′, 5′-GACACCAAGGAAGCCTTAG-3′, 5′-CTCCCACTGGAACAGGTG-3′ and 5′-GGAACAAATAGCATGGATGAC-3′). Sequences were analyzed on the ABI 3700 DNA analyzer. All residue numbers were taken against the HXB reference sequence.

HLA class I typing. HLA class I typing was performed on extracted genomic DNA by PCR single-strand conformation polymorphism[39].

Epitope optimization and recognition of TW10 variant peptides. We confirmed TW10 (TSTLQEQIAW) as the optimal B57-restricted epitope in C-clade infection as previously described[14], using fresh PBMCs from the C-clade-infected subject A-005M (HLA-B57/7). TSTLEQIGW was confirmed as the optimal epitope for B-clade infection in the same way, using PBMCs from the B-clade-infected subject SWS (HLA-B57/13). In both cases, we tested fresh PBMCs for recognition of synthetic TW10 peptides in ELISpot assays as previously described[40].

Selection analysis and phylogenetic inference. A measure of the selection pressures operating at individual codons in the Gag protein was obtained using a maximum-likelihood method[20], which was recently shown to be a powerful way to determine selection pressures in HIV[41]. This method works

by assessing the fit to the data of various models of codon evolution, which differ in how d_N/d_S varies across the sequence and takes into account the phylogenetic relationships of the sequences in question[20]. Two models of codon evolution were used: M7, which specifies ten categories of d_N/d_S in the alignment, none of which may be <1 so that evolution is entirely neutral; and M8, which only differs from M7 in that it incorporates an extra (eleventh) class of codons that can take on any value of d_N/d_S, including those supporting positive selection ($d_N/d_S > 1$). Because M7 and M8 are nested, they may be compared using a standard likelihood ratio test. Consequently, if model M8 significantly rejects model M7 and includes a class of codons where $d_N/d_S > 1$, we can conclude that positive selection has acted. Individual positively selected amino acid sites can then be identified using a Bayesian approach. All analyses were undertaken using the CODEML program from the PAML package[42], with input phylogenetic trees inferred using the maximum-likelihood method available in the PAUP* package[43]. All other phylogenetic trees were also inferred using the maximum-likelihood method available in PAUP*, assuming the GTR + Γ + I model of nucleotide substitution (all parameter values are available from the authors on request). To assess the support for individual nodes on the tree, a bootstrap resampling analysis was undertaken, using the same substitution model but with 1,000 replicate neighbor-joining trees.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

We thank those who made possible the collection of blood samples for this study, in particular the staff at Cato Manor Clinic, Durban (the largest single cohort). We also thank M. Hammond and M. Bunce for HLA typing; C. Hull-Jackson, B. Matz, C. Edwards and H. Coovadia for facilitating these studies; M. Maiden and L. Richardson for assisting with the sequencing; and P. Klenerman and R. Phillips for critically reviewing the manuscript. This work was supported by the Wellcome Trust (P.J.R.G. and A.L.), the Elizabeth Glaser Pediatric AIDS Foundation (P.J.R.G. and M.F.), the Spanish Health Department (BEFI, Red G03/173 to J.M.-P.), the National Institutes of Health (AI46995-01A1 and N01-AI-15442) and the Doris Duke Charitable Foundation. P.J.R.G. is an Elizabeth Glaser Scientist and B.D.W. is a Doris Duke Distinguished Clinical Science Professor.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 November 2003; accepted 8 January 2004
Published online at http://www.nature.com/naturemedicine/

1. Koenig, S. *et al.* Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. *Nat. Med.* **1**, 330–336 (1995).
2. Borrow, P. *et al.* Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* **3**, 205–211 (1997).
3. Goulder, P.J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**, 212–217 (1997).
4. Barouch, D.H. *et al.* Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**, 335–339 (2002).
5. Klenerman, P., Wu, Y. & Phillips, R. HIV: current opinion in escapology. *Curr. Opin. Microbiol.* **5**, 408–413 (2002).
6. Evans, D.T. *et al.* Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nat. Med.* **5**, 1270–1276 (1999).
7. Allen, T.M. *et al.* Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. *Nature* **407**, 386–390 (2000).
8. Phillips, R.E. *et al.* Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**, 453–459 (1991).
9. Moore, C.B. *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439–1443 (2002).
10. Goulder, P.J. *et al.* Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412**, 334–338 (2001).
11. Kaslow, R.A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nat. Med.* **2**, 405–411 (1996).
12. Migueles, S.A. *et al.* HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc. Natl. Acad. Sci. USA* **97**, 2709–2714 (2000).
13. O'Brien, S.J., Gao, X. & Carrington, M. HLA and AIDS: a cautionary tale. *Trends Mol. Med.* **7**, 379–381 (2001).
14. Goulder, P.J. *et al.* Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Res. Hum. Retroviruses* **12**, 1691–1698 (1996).
15. Tang, J. *et al.* Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. *J. Virol.* **76**, 8276–8284 (2002).
16. Yu, X.G. *et al.* Consistent patterns in the development and immunodominance of human immunodeficiency virus type 1 (HIV-1)-specific $CD8^+$ T-cell responses following acute HIV-1 infection. *J. Virol.* **76**, 8690–8701 (2002).
17. Novitsky, V. *et al.* Magnitude and frequency of cytotoxic T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. *J. Virol.* **76**, 10155–10168 (2002).
18. Addo, M.M. *et al.* Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* **77**, 2081–2092 (2003).
19. Altfeld, M.E.A. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17**, 2581–2591 (2003).
20. Yang, Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* **51**, 423–432 (2000).
21. McMichael, A. & Klenerman, P. HIV/AIDS. HLA leaves its footprints on HIV. *Science* **296**, 1410–1411 (2002).
22. Frahm, N. *et al.* Identification of highly immunodominant regions in HIV by comprehensive CTL screening of ethnically diverse populations. *J. Virol.* (in the press).
23. Barber, L.D. *et al.* Polymorphism in the α1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J. Immunol.* **158**, 1660–1669 (1997).
24. Kelleher, A.D. *et al.* Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J. Exp. Med.* **193**, 375–386 (2001).
25. Richman, D.D., Wrin, T., Little, S.J. & Petropoulos, C.J. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc. Natl. Acad. Sci. USA* **100**, 4144–4149 (2003).
26. Cohen, J. HIV. Escape artist par excellence. *Science* **299**, 1505–1508 (2003).
27. Migueles, S.A. *et al.* The differential ability of HLA B^*5701^+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. *J. Virol.* **77**, 6889–6898 (2003).
28. Coffin, J.M. HIV population dynamics *in vivo*: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483–489 (1995).
29. Greene, W.C. & Peterlin, B.M. Charting HIV's remarkable voyage through the cell: basic science as a passport to future therapy. *Nat. Med.* **8**, 673–680 (2002).
30. Novitsky, V. *et al.* Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *J. Virol.* **76**, 5435–5451 (2002).
31. Singh, A.R., Hill, R.L. & Lingappa, J.R. Effect of mutations in Gag on assembly of immature human immunodeficiency virus type 1 capsids in a cell-free system. *Virology* **279**, 257–270 (2001).
32. Furuta, R.A. *et al.* HIV-1 capsid mutants inhibit the replication of wild-type virus at both early and late infection phases. *FEBS Lett.* **415**, 231–234 (1997).
33. Krogstad, P. *et al.* Human immunodeficiency virus nucleocapsid protein polymorphisms modulate the infectivity of RNA packaging mutants. *Virology* **294**, 282–288 (2002).
34. Forshey, B.M., von Schwedler, U., Sundquist, W.I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
35. Braaten, D., Franke, E.K. & Luban, J. Cyclophilin A is required for the replication of group M human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus SIV(CPZ)GAB but not group O HIV-1 or other primate immunodeficiency viruses. *J. Virol.* **70**, 4220–4227 (1996).
36. Braaten, D., Franke, E.K. & Luban, J. Cyclophilin A is required for an early step in the life cycle of human immunodeficiency virus type 1 before the initiation of reverse transcription. *J. Virol.* **70**, 3551–3560 (1996).
37. Yoo, S. *et al.* Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.* **269**, 780–795 (1997).
38. Goulder, P.J. *et al.* Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. *J. Exp. Med.* **193**, 181–194 (2001).
39. Bunce, M. *et al.* Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* **46**, 355–367 (1995).
40. Goulder, P.J. *et al.* Rapid definition of five novel HLA-A*3002-restricted human immunodeficiency virus-specific cytotoxic T-lymphocyte epitopes by Elispot and intracellular cytokine staining assays. *J. Virol.* **75**, 1339–1347 (2001).
41. Ross, H.A. & Rodrigo, A.G. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J. Virol.* **76**, 11715–11720 (2002).
42. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
43. Swofford, D.L. in PAUP*. *Phylogenetic Analysis Using Parsimony (*and other methods)* 4th edn. (Sinauer Associates, Sunderland, Massachusetts, 2002).
44. Hayman, A. *et al.* Phylogenetic analysis of multiple heterosexual transmission events involving subtype B of HIV type 1. *AIDS Res. Hum. Retroviruses* **17**, 689–695 (2001).
45. Kasper, P. *et al.* The genetic diversification of the HIV type 1 gag p17 gene in patients infected from a common source. *AIDS Res. Hum. Retroviruses* **11**, 1197–1201 (1995).